图书情报工作
第 64 卷 第 15 期 2020 年 8 月

新型冠状病毒突发公共卫生事件中的数据共享机制研究

■ 崔宇红 王飒

北京理工大学图书馆 北京 100081

摘 要：[目的/意义]新型冠状病毒疫情下，建立完善数据开放平台和共享机制，应对全球公共卫生挑战，推进公共卫生紧急事件数据共享能力的建设，已经成为各国政府和科技界的普遍共识。[方法/过程]首先审视历次突发公共卫生事件中的数据共享政策规范和应对措施，接着通过对同行评议期刊、预印本、数据知识库、临床试验注册平台和基因组/结构数据中心等数据共享来源的实证调查，分析新型冠状病毒数据的共享现状，最后探讨突发公共卫生事件中不同数据共享机制的优缺点和障碍。[结果/结论]新型冠状病毒疫情的数据共享程度明显提高，但仍未达到常态化。技术、动机、经济、政治、法律和伦理因素是影响数据共享的主要障碍。我国应从战略规划、基础设施、利益相关者和伦理法律等方面，加强突发公共卫生事件中的数据共享能力。

关键词：数据共享机制 突发公共卫生事件 新型冠状病毒

分类号：G253

DOI：10.13266/j.issn.0252－3116.2020.15.013

2020 年 1 月 30 日，世界卫生组织（简称"世卫组织"）宣布将新型冠状病毒（简称"新冠病毒"）疫情列为国际关注的突发公共卫生事件，在突发公共卫生事件中，共享数据对于制定有效的公共卫生应急管理策略至关重要。数据是公共卫生行动的基础，成功的公共卫生数据共享政策措施有助于疾病的早期预防，及时与全球共享、沟通临床试验数据和分子流行病学信息，采取适当的干预措施和新的行动以避免出现大规模病患或导致死亡等灾难性后果。近年来，多起全球突发公共卫生事件（如 2009 年的甲型流感、2013－2016 年西非埃博拉病毒病疫情、2015 年寨卡病毒综合症）引发对公共卫生领域关于研究数据共享问题的广泛讨论。全球传染病防备研究合作组织（Global Research Collaboration for Infectious Disease Preparedness，GloPID-R）在引发传染病大流行的 12 种病原体文献、数据和政策搜索基础上，对常用的数据共享方式和平台进行分类调查，以评估潜在大流行病原体的数据共享实践[1]。为了加深对紧急情况下数据共享的障碍和影响因素的了解，惠康基金和英国国际发展部委托 GloPID-R 数据共享工作组编写 6 个案例研究，强调不同流行病暴发场景下数据共享原则在实施中需要考虑的复杂性[2]。据《自然》（*Nature*）杂志报道，在新型冠状病毒疫情暴发的一个月内，科研人员已发表至少 80 篇论文[3]，而笔者截至 2020 年 5 月的统计显示，新冠病毒研究论文数量呈指数增长趋势，新冠病毒的全球快速传播也更加凸显了公共卫生领域数据共享实践面临的挑战[4]。与此同时，各基金组织和出版机构迅速做出响应，通过开放获取提供对新冠病毒研究文章的公开访问，鉴于此，本研究旨在通过对重大突发公共卫生事件中数据共享政策的梳理和新冠病毒数据共享实践状况的调查，探讨突发公共卫生事件中不同数据共享机制的特点和障碍，旨在推进我国公共卫生紧急事件数据共享能力的建设。

1 重大突发公共卫生事件中数据共享的规范与举措

20 多年来，新发和再发病毒感染和大流行疫情对全球公共卫生安全构成重大威胁。按照疾病暴发时间的先后，本文选择引起公众广泛关注的几次重大疫情，即 1997 年中国和许多亚洲国家的 H7N9 流感以及 2009 年席卷全世界的 H1N1 流感，2013－2016 年拉病毒病疫情和 2015 年寨卡病毒综合症，以及 2019 年底暴发的新冠病毒疫情，可以看到，世界卫生组织、各国基金组织和顶级出版机构迅速对疫情做出响应，极大

作者简介：崔宇红（ORCID：0000－0002－5215－7726），副馆长，研究馆员，博士；E-mail：cuiyh@bit.edu.cn；王飒（ORCID：0000－0003－4665－5419），副研究馆员，博士。

收稿日期：2020－03－25 修回日期：2020－05－29 本文起止页码：104－111 本文责任编辑：王传清

地促进了全球数据共享倡议组织、数据共享国际规范、期刊出版政策指南的建立、完善和实施。

1.1　全球共享流感数据倡议组织

1997 年,人类可以感染禽流感及其他人畜共患型流感病毒,如甲型 H5N1、甲型 H7N9 和甲型 H9N2 等禽流感病毒亚型,引起人们对可能暴发流感大流行的强烈关注。令人震惊的是,当第一批 H5N1 病例在印度尼西亚被发现时,世卫组织几乎没有可利用的最新动物序列数据来帮助其了解病毒的进化。论文发表的优先权之争导致科研人员通常不愿公开新兴突发疾病的原始材料和数据,即使是由美国疾控中心资助建立的 H5N1 数据库也只允许提供 H5N1 序列数据的国家使用。为了解决在发表后通过公共领域数据库共享数据的棘手问题,2006 年,70 多位著名科学家在《自然》杂志上联名呼吁成立全球禽流感数据共享倡议组织(Global Initiative of Sharing All Influenza Data, GISAID),倡议要求所有数据在提交后的 6 个月内要发布到 GenBank 和其他公共领域档案库中,所有信息获取者都必须遵守知识产权和信用归属规则[5]。

在大流行性病毒出现的情况下,传统的公共领域档案库已经被证明不能成功地实现重要数据的快速共享,GISAID 为快速共享已发布和未发布的流感数据提供一种有效且可信赖机制。2008 年,GISAID 托管的 EpiFlu 数据库启动,各国能够随时随地跟踪新病毒在全球传播和演变,这在应对 2009 年甲型 H1N1 流感暴发时意义重大。2013 年,中国科学家率先发布甲型 H7N9 病毒基因序列,使用此序列数据在几个星期内合成生物学试验开发和测试疫苗病毒,进一步说明数据提供者和用户对 GISAID 共享机制的接受和信任,证明 GISAID 在及时共享重要流感数据方面的可行性[6]。2020 年 3 月,深圳国家基因库与 GISAID 达成战略性合作,深圳国家基因库生命大数据平台成为 GISAID 在中国的首个正式授权平台。

1.2　数据共享国际规范

2014－2016 年西非埃博拉病毒病的大规模暴发,2015 年寨卡病毒感染引发的格林－巴利综合症和婴儿小头症,极大地推动了数据共享国际规范的出台和相关措施的完善。2015 年 9 月,世卫组织召开磋商会议,就公共卫生紧急情况下如何尽可能实时公开地提供数据和研究成果制定全球规范,针对担忧出版前披露关键信息等问题,明确在公共卫生应急事件发生情况下,期刊应该鼓励或要求在文章发表前共享数据,包括实验室研究结果(如基因组和免疫学数据)、人口统计学数据、动物研究结果以及研究参与者的临床数据[7]。

为支持世卫组织全球数据共享规范的实施,2016 年 2 月,英国医学科学院、美国国家卫生研究院、美国自然科学基金会、惠康基金会、比尔和梅林达·盖茨基金会、生物技术与生物科学研究理事会以及《英国医学杂志》(*BMJ*)、《柳叶刀》(*The Lancet*)、《科学》(*Science*)、PLOS 等 31 家机构共同签署声明,承诺完善当前以及未来突发公共卫生数据开放机制,减少共享科研成果的延迟。该声明强调公共卫生紧急情况下数据共享的实施路径,即期刊承诺“免费获取寨卡病毒相关的所有内容。任何为不受限制的传播而早于论文提交的数据和预印本不会在这些期刊上抢先发表”,资助者“要求从事与公共卫生紧急情况有关的研究人员建立适当的机制,以尽快广泛地与公共卫生和研究团体以及世卫组织共享有质量保证的中期和最终数据”[8]。

国际医学期刊编辑委员会(International Committee of Medical Jouranl Editors,ICMJE)于 2016 年 1 月 20 日发布公告,要求 2018 年 7 月及以后提交到 ICMJE 期刊的临床试验报告,必须包含数据共享声明,2019 年 1 月以后入组受试者的临床试验必须在临床试验注册平台上提交数据共享计划。如果数据共享计划有变化或更改,应在注册平台上进行更新并在提交论文时加以说明[9]。

由全球健康安全中心领导、比尔和梅琳达·盖茨基金会支持的“增强公共卫生数据共享”项目组织了一系列圆桌会议,交流如何创建正确的数据共享环境、实践案例和经验。为解决最佳数据共享的政策和技术问题,项目开发了“共享公共卫生监测数据和收益指南”,指南主要针对全球卫生领域的合作伙伴,包括公共卫生部门、非政府组织、私营部门、学术机构、多边组织、出版商、基金组织和其他机构,旨在推动将数据共享国际规范成为尽可能公开和适当的数据共享模型[10]。

1.3　新冠病毒疫情暴发以来的政策响应

2019 年 12 月以来,新冠病毒的迅速传播对全球健康构成重大威胁。为解决疾病暴发情况下提交论文的数据访问障碍问题,《世界卫生组织简报》实施“COVID-19 Open”数据共享和报告机制[11],即当向《简报》提交数据后,所有与新冠病毒疫情有关的研究论文都将被分配一个数字对象标识符,并在接受同行评议的 24 小时内在线发布到“nCov-2019 Open”数据平台上。在遵守知识共享协议的前提下,明确论文数据的所有权归属论文作者,数据可以在任何媒体上不受限制地

自由使用、分发和复制,前提是必须按照知识共享署名 3.0 协议的要求适当引用原始作品。如果论文经同行评议被录用,其开放评议结果也将在最后发表的出版物中报告。如果论文不能通过同行评议,作者可以选择向其他期刊投稿和出版,或者发表在世卫组织推荐的其他快速共享论文和数据的平台上,如 PROMED 和 F1000Rsearch。

惠康基金会、自然等出版商、基金资助机构和科技社团组织签署联合声明,以确保与冠状病毒相关的研究数据和发现的快速共享[12]。声明强调的几点原则是:①所有与疾病暴发有关的同行评审论文均应立即开放获取。②与疾病暴发有关的研究发现在提交到期刊后,在期刊和作者知情情况下应立即与世卫组织共享。③期刊在正式发表或者同行评议前,可通过预印本服务等开放获取平台来获得研究结果。④研究人员、公共卫生研究机构以及世卫组织应尽快、广泛地共享与疫情相关的研究数据,以及数据采集的协议和标准。⑤作者提交的共享数据或预印本不会抢先被期刊发表。

2020 年 2 月 11 日至 12 日,在世卫组织召开的"新冠肺炎全球研究与创新论坛"上,数据共享被列入 8 项优先研究议程之一,来自世界各地的科学家一致认为应迅速分享病毒资源、临床样本和相关数据,以用于当前的公共卫生实践,同时用这些资源开发的医学产品或创新成果也必须作为共享的一部分,公平公正地获取[13]。

由此可见,国际组织和出版机构已经采取积极的行动来应对新冠病毒的数据共享需求和挑战,下面将针对这些政策和原则是否已经以及如何转化为实践和现实开展实证研究。

2 新冠病毒突发公共事件中的数据共享状况调查

为了解新冠病毒疫情暴发以来数据共享的现状并与以往研究进行比较,选择同行评议期刊、预印本、数据知识库、临床试验注册库和基因组/结构数据中心 5 种数据共享来源,检索不同方式下的新冠病毒论文关联数据、临床试验记录和基因组数据的开放程度。检索词设置为"coronavirus"或"2019-nCov"或"COVID-19",时间范围限定在 2020 年 1 月 1 日至 5 月 29 日期间。

2.1 同行评议期刊

从科学研究的规范性角度来看,同行评议期刊的评议制度有效保证了高质量研究成果的共享传播。选择 Elsevier、Wiley、Springer、Taylor & Francis 四大出版商平台和《柳叶刀》、《医学病毒学杂志》(*J MED VIROL*)、《自然》、《新英格兰医学杂志》(*NEJM*)、《中华结核和呼吸杂志》5 种专业期刊,检索并人工筛选获得新冠病毒密切相关文献(论文类型仅统计"article"和"review"),然后检查每篇论文是否有数据共享声明和补充材料,检索结果见表 1 和表 2。如果有共享数据,则这些数据应该关联到所发表的论文,对于提供基础数据作为补充材料,或提供访问基础数据声明的论文均被视为具有数据共享的论文。该定义也适用于下面的预印本。

几乎所有新冠病毒期刊论文均可以通过开放获取(open access)或免费全文获取(free text access)方式不受限地访问。如表 1 所示,四大出版商平台近 20% 的新冠病毒相关论文有共享数据,但不同出版商平台在共享数据方式和程度上存在差异。Elsevier 发表的新冠病毒论文数量和论文数据共享量均为最高,平台提供数据声明、数据获得性和附件材料来共享数据,如数据声明中指出由研究人员根据使用者的申请来确定是否共享数据,或者给出论文中所用数据的来源和可访问性,作为附件材料下载共享的原始数据格式包括 PDF、表格、视频、音频、图像、软件代码等。

表 1 出版商平台发表新冠病毒论文的数据共享情况

项目	Elsevier	Wiley	Springer	Taylor & Francis	总计
检索密切相关论文数(篇)	2 085	329	323	297	2 994
有共享数据的论文数(篇)	434	28	95	25	582
共享数据论文的百分比(%)	20.8	8.5	29.4	8.4	19.4

与出版商平台相比,不同期刊在新冠病毒数据共享上的差异性更加显著。如表 2 所示,*The Lancet*、*Nature* 和 *NEJM* 的数据共享情况显著高于出版商平台和其他期刊,而《中华结核和呼吸杂志》官网上检索到的新冠病毒论文则没有提供任何数据共享方式。

2.2 预印本

为了加快新冠病毒研究成果的广泛传播,研究人员在正式投稿或者被期刊所接纳前,可以将科研论文

表2 专业期刊发表新冠病毒论文的数据共享情况

项目	*The Lancet*	*J MED VIROL*	*Nature*	*NEJM*	《中华结核和呼吸杂志》
检索密切相关论文数(篇)	26	101	23	13	9
有共享数据的论文数(篇)	22	16	13	13	0
共享数据论文的百分比(%)	84.6	15.8	56.5	100	0

的手稿提交到预印本平台。需要强调的是,这些未经同行评议的初步研究不应被视为结论性报告用于指导临床实践和健康行为,也不应作为既定事实在新闻媒体上报道。一个典型例子是,1月31日印度理工学院的研究人员在bioRxiv发表的新冠病毒论文被广泛转载但之后遭到学术界的强烈质疑,认为有夸大和扭曲事实的嫌疑,最终作者撤稿[14]。

选择arXiv、SSRN、bioRxiv、chemRxiv、medRxiv和ChinaRxiv等6个预印本平台检索并人工筛选获得新冠病毒密切相关文献,然后检查每个平台是否提供数据共享功能,如果提供则计算有共享数据论文的比例情况,如表3所示:

表3 预印本的数据共享情况

项目	arXiv	SSRN	bioRxiv	chemRxiv	medRxiv	ChinaXiv
检索密切相关论文数(篇)	1 105	17	1 045	414	3 280	29
有共享数据的论文数(篇)	0	0	546	222	3 260	3
共享数据论文的百分比(%)	0	0	52.2	53.6	99.4	10.3

由表3可见,预印本平台上新冠病毒相关论文的数据共享实践与所隶属的学科领域显著相关,生命健康领域预印本在新冠病毒论文数据共享方面发挥主要作用,而两个历史悠久的预印本服务,主要是面向物理学和数学领域的arXiv与主要面向社会科学和人文学领域的SSRN均没有提供附件文档分享数据的功能。SSRN在2016年被Elservier收购,从2018年起与*The Lancet*和*Cell*合作,作者在将稿件提交到*The Lancet*及旗下期刊时,可以选择同意将预印本发布在SSRN以快速共享研究成果,但SSRN也没有对论文数据提供任何形式的访问。

bioRxiv、chemRxiv和medRxiv是近年来新推出的生命健康领域预印本服务,3个预印本平台的论文数据共享的比例远高于同行评议期刊。medRxiv是bioRxiv平台的衍生产品,两者在数据共享的实现方式上都提供补充材料和数据/代码两种途径。补充材料可以从预印本平台直接获得多种格式的数据文件,数据/代码则指明数据的可获得性,或者提供到外部数据存储平台(如量化生物医学研究中心、GISAID和Github等)的链接来共享数据。相比之下,medRxiv所有文章都在数据/代码中提供数据共享声明,但仅有27篇论文同时以补充材料来提供数据,而bioRxiv则主要以补充材料方式共享数据。chemRxiv采用Figshare平台同时支持论文和数据文件的上传和关联。3个平台都自动分配数字对象标识符(DOI),整合来自Altmetric的在线社交媒体提及的内容和指标,以实现对新冠病毒相关信息的早期评估和加速传播。

自疫情暴发以来,中国科学院科技论文预发布平台ChinaXiv全力支持国内外新冠病毒相关研究内容的快速发表,在平台上发表的论文可以附件形式下载数据。

2.3 数据知识库

如果研究人员愿意发布数据,最直接的方式就是构建更多专门用于数据存储和共享的知识基础设施,近些年刚刚起步的数据知识库能够保障数据在最大程度上被可靠地获取和使用。基于DataCite检索平台,选择Zenodo、Figshare Academic Research System(Figshare ARS)、Mendeley Data、Harvard Dataverse等4个数据知识库以考察新冠病毒的数据共享程度和平台数据管理功能,见表4。Zenodo是根据欧洲OpenAIRE计划开发并由CERN运营的通用开放访问存储库。Zenodo将软件代码存储在Github平台上。Figshare ARS与chemRxiv、Taylor & Francies、Frontiers、Mendely等建立数据共享服务合作关系,许多期刊或预印本平台支持将论文的补充数据存储在Figshare ARS,从Figshare ARS也可以关联到同行评议出版物[19]。Harvard Dataverse是哈佛大学专为存储和共享大学研究数据的机构知识库,数据集主要来自哈佛大学中国数据实验室,包括:COVID病例地图,百度流动数据,病例更新,卫生疗疗机构设施,病毒报告,中国政策与舆情信息搜集,地理信息系统与公共卫生下的研究论文和综述。

4个数据知识库平台均支持数字对象唯一标识符DOI,这是发现、保存和引用研究数据的关键。相比于Harvard Dataverse复杂的元数据元素和标准,Figshare

ARS 和 Zenodo 采用的是极简原则的元数据标准，便于快速方便地存储研究数据。此外，Figshare ARS 和 Zenodo 均提供 altmetrics 指标，支持多种类型的文件上传和下载，包括文本、图像、数据集、视音频、软件、交互资源等。

表 4 数据知识库的数据共享情况 （单位：项）

名称	文本	图像	数据集	软件	交互资源	音视频	馆藏	其他
Zenodo	1 072	104	569	235	9	26	5	71
Figshare ARS	556	217	499	30		98	88	166
Mendeley Data			349					
Havard Dataverse			2 076					

2.4 临床试验注册平台

出于伦理和科学两方面的需求，所有临床试验均应注册。由世卫组织牵头建立的国际临床试验注册平台（International Clinical TrialsRegistry Platform，ICTRP），对符合条件的各国注册机构进行认证，是全球临床试验一站式检索入口。如表 5 所示，在 ICTRP 搜索共发现 2 936 项新冠病毒临床试验，按照提供临床试验记录的注册来源地统计，排名前 5 的机构分别是美国临床试验数据库（ClinicalTrials. gov）、中国临床试验注册中心（Chinese Clinical Trial Registry，ChiCTR）、伊朗临床试验注册中心（Iranian Registry of Clinical Trials，IRCT）、欧盟临床试验注册中心（EU Clinical Trials Register，EU CTR）和德国临床试验注册中心（German Clinical Trials Register，GermanCTR），这表明在新冠病毒疫情公共卫生紧急情况下，中国临床试验注册机构与国际同行一起，正在广泛开展和推动快速共享数据。

从 2005 年开始，ICMJE 采取了一项鼓励临床试验前瞻性注册的政策，要求临床试验必须在招募第一个患者时或招募前（即“前瞻性”）进行登记，并作为在 ICMJE 成员期刊上发表研究的先决条件[15]。结果显示，在新冠病毒已招募研究对象的 1 443 项临床实验中，有 686 项是前瞻性注册，这说明尽管大部分期刊都参照 ICMJE 规则制定稿件发表指南，但通常并没有严格坚持遵循前瞻性注册的政策。

表 5 临床试验注册机构的数据共享情况

项目	美国	中国	伊朗	欧盟	德国
临床试验记录数（项）	1 590	688	171	161	69

2.5 基因组/结构数据中心

通常针对某种病原体的遗传基因序列和蛋白质结构等信息存储在特定的大型国际级数据中心，对于开发用于诊断的检测试剂，追踪疾病的持续暴发以及选择潜在的干预措施至关重要。为及时向全球公众提供新冠病毒的相关信息，中国国家生物信息中心（CNCB）/国家基因组科学数据中心（NGDC）建立了 2019 新冠病毒信息库（2019nCoVR）[16]，整合来自世界主要数据中心公开发布的 2019-nCoV 核苷酸和蛋白质序列数据，包括德国的全球流感病毒数据库（GISAID）、美国国家生物技术信息中心的 GenBank、国家基因库生命大数据平台（CNGBDB）、国家微生物科学数据中心和国家病原微生物资源库联合建设的新冠病毒国家科技资源服务系统（NMDC）以及 CNCB/NGDC 的 Genome Warehouse 等，见表 6，超过 3 万种冠状病毒序列涉及 541 个数据递交单位、598 个样本采集单位和 456 个采集地点。

全球越来越多的实验室通过 GISAID 发布新冠病毒的基因组序列。科研工作者在 GISAID 注册后都可上传他们提取的病毒基因序列，每个毒株都会有个独一无二的编号，采集时间、提交日期、提交实验室等信息也都记录在案。按照样本采集时间，最早的一条新冠病毒基因序列是在 2019 年 12 月 23 日采集的，并由中国医学科学院病原所于 1 月 11 日分别上传到 GISAID 和 Genome Warehouse 中。这些基因序列数据通过同行评审医学期刊可以进行迅速的信息传播，以便在流行病发生时将相关信息在线发布。例如文章[17]从分析的 9 位患者的样本中，获得了 2019-nCoV 的 8 条完整和 2 条部分基因组序列。这些数据已保存在中国微生物数据中心（登录号 NMDC10013002 和基因组登录号 NMDC60013002-01 至 NMDC60013002-10），而通过华大基因采集的数据已保存在中国国家基因库（登录号 CNA0007332-35）。

3 突发公共卫生事件数据共享机制的比较和障碍分析

如果将出版物比作科学宇宙中的恒星，那么数据就是学术交流中的“暗物质”[18]。近年来鼓励更加开放的数据共享政策推动数据共享的实现途径更加多元化，与以往的大流行疫情或突发公共卫生事件相比，新冠病毒的数据共享程度有了明显提高，但从同行评议

表 6　基因组/结构数据中心的数据共享情况

项目	GISAID	GenBank	Genome Warehouse	NMDC	CNGBdb
新冠病毒基因组序列数据量(项)	30 029	4 669	32	146	75

期刊、预印本和数据知识库的论文数据共享比例来看，按照“开放为常态、不开放为例外”的共享理念，显然距离数据共享常态化的目标还有相当的距离。

总体来看，现有的数据共享机制还达不到开放科学的两个目标——透明性和实用性，特别是在如何融合不同来源的数据以更快地产生新知识和制定决策方面。表 7 分析了 5 种数据共享机制的优缺点，下面主要基于快速共享、可重用、可持续这 3 个数据共享最重要的属性，对不同数据共享机制进行分析。

表 7　公共卫生紧急事件中数据共享机制的比较

共享机制	优点	缺点
同行评议期刊	同行评议保障论文和数据的质量和可靠性；出版商和期刊的支持使得很容易发现和获取相关数据	发表速度慢；PDF 等文件格式无法机读，限制了可发现性和软件智能分析
预印本	生命健康领域预印本被广泛接受；发表速度快于同行评议期刊	论文和数据的可靠性未经证实；PDF 等文件格式无法机读，限制了可发现性和软件智能分析
数据知识库	潜在的高成本效益；平台可能支持更丰富的数据管理功能	标准化元数据很少，数据不容易发现；平台在数据长期保存和可持续性上还有待证明
临床试验注册平台	专业领域元数据标准化，促进了重用和互操作性	数据不能完全满足期刊政策的规范性要求
基因组/结构数据中心	专业领域元数据标准化，保障数据可发现性、完整性和互操作性	可能存在大量的重复劳动；昂贵的平台建设和维护成本

发表论文是共享研究成果最常见的形式，通过论文发现和关联数据对保证研究完整性非常重要。尽管同行评议过程是质量控制的有效手段，但在大流行暴发阶段很难满足快速分享的需求。预印本旨在弥补期刊出版时间较长的缺陷，但却面临无法保证质量、没有被主流搜索引擎(如 Web of Knowledge，Scopus 等)索引以及研究人员担心预印本共享论文和数据会影响在高影响力期刊发表等诸多困扰，这也凸显了数据共享速度和质量保证之间的矛盾。此外，这两种共享机制面临的共同问题是基于传统出版模式下的数据文件很难被发现和重用。2020 年 3 月 13 日，美国和其他 11 个国家/地区的政府科学顾问呼吁出版商在通过预印本、知识库或其他来源共享数据的同时，还要提供可被人工智能软件和计算机系统读取的电子表格等格式的数据，而不仅仅是 PDF 文档。

与前两种数据共享方式一样，数据知识库可以通过共享论文数据来提高研究透明度，但在解决数据重用这一更高层次的目标上仍面临困难。随着全球数据生态系统向学科无关的通用型知识库转移，大多数数据知识库使用最小要素的元数据标准(如 Dublin Core)，方便存储数据和实现与其他数据源的互操作。本研究中，Figshare ARS 共享的新冠病毒数据集最多，使用 DataCite 的元数据标准，可以自动检查数据完整性，遵守创作共享许可协议并寻求与学术出版商和机构的合作。数据知识库的资金来源较为复杂，Figshare ARS 是由霍尔茨布林克出版集团提供资金支持，Mendeley 则被爱思维尔收购成为其旗下产品，同时 Mendeley 将所有数据备份到 DANS，这是荷兰政府支持的数据归档服务，而 Harvard Dataverse 依靠哈佛大学的资金支持，从数据长期保存角度来看，数据知识库平台的可持续性还需要时间的证明。

在强大的专业社区支持下，特定病原体的临床试验注册平台和基因组/结构数据中心的数据标准化程度最高，最易于重用。这些国际级或国家级的数据共享基础设施，无疑需要投入巨大的资金和人力成本。如美国国家医学图书馆 2016 年预算拨款中，有 1.9 亿美元用于处理并提供新型测序、微阵列和小分子筛选技术的大量数据以及临床试验数据，美国国家生物技术信息中心有 288 名全职员工负责 GenBank 数据库的管理和维护工作。数据中心昂贵的运行费用使得绝大多数基因组测序工作是在高收入国家进行的，因此可能无法保障低收入国家研究人员的利益。

综上所述，公共卫生应急事件中快速数据共享的主要障碍包括：①缺乏适当的数据共享基础设施，如数据中心和知识库平台；②数据提供者和数据用户在数据归属和学术认可上缺乏明确的动力机制，紧急情况对研究人员共享数据行为的激励有限；③研究成果商业化和知识产权可能带来的经济利益损失；④高收入和低收入国家和地区研究人员在能力和资金方面的不平等，不同研究团体或机构使用数据管理系统的费用和程度不同；⑤过早分享数据和研究成果可能带来学术声誉风险；⑥实验治疗和临床护理中与隐私和知情权有关的监管和治理问题，以及遵守国家和国际道德和法律要求。这些障碍对应技术、动机、经济、政治、法律和伦理 6 类因素，其中前 3 项最为重要，不同因素之

间复杂的相互关联和作用反馈，使得任何针对单一因素的解决方案都很难取得满意的结果。

4 推动我国突发公共卫生事件数据共享能力建设的建议

我国在新冠病毒疫情暴发初期，率先向世界发布病毒基因组序列数据和相关信息，展现了中国在应对突发公共卫生事件中的公开透明和负责任的态度，但也出现了科研人员因考虑论文优先发表权而不及时披露信息，或者因提前公开数据被其他研究者利用数据抢先发表论文的事件。在现实世界中，全球公共卫生领域数据共享框架的建立和完善是从危机中衍生，而不是来自前瞻性的规划。考虑到社会、政治和文化背景，特别是各国法律制度体系的巨大差异，强制性、“一刀切”的全球数据共享机制在实践中往往效果不佳。从历史经验来说，应在遵循全球数据共享原则框架基础上，基于本地需求和治理特征，从战略规划、基础设施、利益相关者、伦理及法律等方面，提升我国在突发公共卫生事件中的数据共享能力。具体建议如下：

4.1 明确数据共享在国家突发公共卫生事件防范和应对中的战略地位

2018 年 4 月，我国国务院办公厅出台科学数据管理办法，首次从国家层面上制定提高数据开放共享水平的政策原则。在世卫组织发布的新冠病毒研发路线图中，数据共享被列为须优先关注的核心任务之一，主要涉及评估需求、制定目标、协调实施有针对性的干预措施以及对干预活动进行监测和评估。具体来说，我国应在公共卫生领域数据收集目的、用户需求以及参与数据收集、管理、分析和使用的行为者生态系统方面，明确制定我国数据共享的实施路线图，为当前和长期公共卫生所需的信息管理系统和程序提供信息，以便在公共卫生事件或紧急情况发生之前、期间和之后提升决策者和公众的意识并影响其行为。

4.2 发展和建设跨地域、学科和媒体的全球数据共享基础设施

仅仅在原则上就共享原始数据达成国际共识是不够的。科研数据是未来科技创新的“石油”，也是国家人口健康和生物安全的重要基础资源，从上述研究可以发现，当前常用的数据共享平台和权威期刊的所有权仍主要掌握在美国、英国等科技领先国家或国际性研究机构中，要在未来的科技竞争中处于领先位置，我国首先应针对具有流行病潜力的疾病开发数据共享平台，包括创建开放获取期刊、预印本、临床试验注册平台、基因组/结构数据库等，以支持国际国内研究合作。其次，应致力于建立通用型数据知识库平台，明确数据标准化方法，开发标准化工具，简化和统一数据监管流程，以确保数据质量来应对实施数据共享协议的技术挑战。最后，综合利用社交媒体、传统媒体及其他快速共享途径，以实现经核实准确信息的传播和评估。

4.3 组织和协同公共卫生突发事件风险中数据共享利益相关者的行动

参与数据共享的主要利益相关者可以划分为数据提供者、数据使用者和数据共享促进者 3 类群体，负责任的利益相关者应具有共享数据和将数据最终整合到决策中的意愿、技能和能力，要克服数据共享障碍需重点关注不同机构和主体间的协调，这主要包括两个方面：其一，研究人员往往同时是数据提供者和使用者，需要在数据归属和学术认可上制定合理的数据共享激励手段，要为研究人员提供培训和宣传材料，使研究人员了解公共卫生紧急事件中数据共享的益处，以增强国家疫情报告的透明度，促进我国数据共享文化的形成和普及。其二，数据共享促进者涵盖政府和基金组织的政策制定者、期刊和数据库平台提供商、图书馆和大学等。政府、基金组织、大学等管理机构要积极促进承认研究质量和公共卫生影响的资助政策，特别是期刊和数据库平台提供商要转变出版模式和经营策略，激励及时共享数据，而不是优先出版。图书馆要融入到科研人员的研究全过程中，针对不同阶段的数据需求提供共享服务，开拓新的服务方式和服务领域。

4.4 加强数据共享的伦理和法律问题研究

在紧急情况下，以透明的方式分享研究中产生的数据可以提高科学价值，数据生产者和出版商对此负有伦理责任，即在合适的知情程序和保护个人隐私的前提下收集和共享数据，促进公共卫生数据的快速传播。此外，一方面研究人员面临开展原创研究和在高影响因子期刊上发表的压力，另一方面企业赞助药物研究是出于商品驱动的经济利益，因此，需要详细考虑到研究人员、研究赞助商、公共卫生人员和出版商与更广泛的研究界、卫生和政府机构和公众在管理和分享研究数据方面的伦理责任问题。鉴于突发事件的潜在严重性和时间敏感性，为及时制定或实施有效的公共卫生措施，还需要明确共享数据的法律责任。根据世卫组织制定的新冠病毒研发路线图，数据伦理与数据共享是密切相关的跨学科优先研究主题。本文没有在数据伦理和法律方面展开进一步的讨论，这将是今后的研究工作之一。

参考文献：

[1] PISANI E, GHATAURE A, MERSON L. Data sharing in public health emergencies[EB/OL]. [2020 - 03 - 10]. https://www.

glopid-r. org/wp-content/uploads/2017/02/data-sharing-in-public-health-emergencies-case-studies-workshop-reportv2. pdf.
[2] Data sharing in public health emergencies: learning lessons from past outbreaks[EB/OL]. [2020－03－10]. https://www. glopid-r. org/wp-content/uploads/2017/02/data-sharing-in-public-health-emergencies-case-studies-workshop-reportv2. pdf.
[3] ENSERINK M. Calling all coronavirus researchers: keep sharing, stay open[J]. Nature,2020,578(7793): 7.
[4] YOZWIAK N, SCHAFFNER S F, SABETI P C. Data sharing: make outbreak research open access[J]. Nature, 2015, 518 (7540): 477－479.
[5] Boosting access to disease data[J]. Nature, 2006, 442(7106): 957－957.
[6] SHU Y, MCCAULEY J. GISAID: global initiative on sharing all influenza data - from vision to reality[J]. Euro surveill,2017,22 (13):30494.
[7] LITTLER K, BOON W M, CARSON G,et al. Progress in promoting data sharing in public health emergencies[J]. Bull world health organ,2017,95(4):243.
[8] WELLCOME TRUST. Statement on data sharing in public health emergencies[EB/OL]. [2020－03－10]. https://wellcome. ac. uk/ press-release/statement-data-sharing-public-health-emergencies.
[9] TAICHMAN D B, BACKUS J, BAETHGE C, et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors[J]. Annals of internal medicine, 2016, 164 (7): 505－506.
[10] A guide to sharing the data and benefits of public health surveillance [EB/OL]. [2020－03－20]. https://www. chathamhouse. org/ sites/ default/files/publications/research/2017-05-25- data-sharing-guide. pdf.
[11] MOORTHY V, RESTREPO A M H, PREZIOSI M P, et al. Data sharing during the novel coronavirus public health emergency of international concern[J]. Bulletin of the World Health Organization. 2020,98(3):1－3.
[12] Sharing research data and findings relevant to the novel coronavirus (nCoV) outbreak[EB/OL]. [2020－03－01]. https://wellcome. ac. uk/ press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-covid-19-outbreak.
[13] COVID 19 Public Health Emergency of International Concern (PHEIC) global research and innovation forum: towards a research roadmap[EB/OL]. [2020－03－04]. https://www. who. int/ blueprint/priority-diseases/ key-action/Global_Research_Forum_FINAL_VERSION_for_web_14_feb_2020. pdf? ua=1.
[14] PRADHAN P, PANDEY A K, MISHA A, et al. Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag[EB/OL]. [2020－03－01]. https://www. biorxiv. org/ content/10. 1101/2020. 01. 30. 927871v1. full. pdf.
[15] HARDEMAN M. Coronavirus research on figshare[EB/OL]. [2020－03－13]. https://doi. org/10. 6084/m9. figshare. c. 4836594. v8.
[16] GOPAL A D, WALLACH J D, AMINAWUNG J A, et al. Adherence to the International Committee of Medical Journal Editors' (ICMJE) prospective registration policy and implications for outcome integrity: a cross-sectional analysis of trials published in high-impact specialty society journals[J]. Trials, 2018, 19(1): 1－13.
[17] 赵文明，宋述慧，陈梅丽，等. 2019 新型冠状病毒信息库[J]. 遗传,2020,42(2):212－221.
[18] LU R, ZHAO X, LI J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding[J]. The lancet, 2020,395(10224): 565－574.
[19] 伯格曼. 大数据、小数据、无数据:网络世界的数据学术[M]. 孟小峰,张祎,赵尔平,等译. 北京:机械工业出版社, 2017.

作者贡献说明：

崔宇红：研究框架设计、数据分析与论文撰写；
王飒：数据采集与政策建议部分内容撰写。

Research on Data Sharing Mechanisms in Novel Coronavirus Public Health Emergency

Cui Yuhong　Wang Sa
Beijing Institute of Technology Library, Beijing 100081

Abstract: [**Purpose/significance**] Under the novel coronavirus epidemics, it has become the general consensus by government and scientific discipline to establish and improve the open data platform and data sharing mechanism, to response the global public health challenge and to promote the construction of local public health emergency management data sharing capacity. [**Method/process**] First, the data sharing policy and actions in previous public health emergencies were examined. Then, based on the investigation of data sharing practices such as peer-reviewed journals, preprint, data repositories, clinical trial registration platform and genome/structure data center, this paper analyzed the current situation of COVID-19 data sharing. Finally, it discussed the advantages, disadvantages and obstacles of different data sharing mechanisms in public health emergencies. [**Result/conclusion**] The COVID-19 data sharing degree was obviously improved, but it had not reached the norm. Technical, motivational, economic, political, legal and ethical factors were barriers to data sharing. China should enhance the ability of data sharing in public health emergencies from the aspects of strategy, infrastructure, stakeholder and ethics.

Keywords: data sharing mechanisms　public health emergency　COVID-19